(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

(12) Übersetzung der **europäischen Patentschrift**

(87) EP 0 351 848 B1

(10) **DE 689 15 353 T 2**

(51) Int. Cl.[5]: **G 10 L 5/04**

(21) Deutsches Aktenzeichen: 689 15 353.8
(86) Europäisches Aktenzeichen: 89 113 343.1
(86) Europäischer Anmeldetag: 20. 7. 89
(87) Erstveröffentlichung durch das EPA: 24. 1. 90
(87) Veröffentlichungstag der Patenterteilung beim EPA: 18. 5. 94
(47) Veröffentlichungstag im Patentblatt: 20. 10. 94

(30) Unionspriorität: (32) (33) (31)
21.07.88 JP 183906/88

(73) Patentinhaber:
Sharp K.K., Osaka, JP

(74) Vertreter:
Tauchner, P., Dipl.-Chem. Dr.rer.nat.; Heunemann, D., Dipl.-Phys. Dr.rer.nat.; Rauh, P., Dipl.-Chem. Dr.rer.nat.; Hermann, G., Dipl.-Phys. Dr.rer.nat.; Schmidt, J., Dipl.-Ing.; Jaenichen, H., Dipl.-Biol. Dr.rer.nat., Pat.-Anwälte; Tremmel, H., Rechtsanw., 81675 München

(84) Benannte Vertragstaaten:
DE, FR

(72) Erfinder:
Kitoh, Atsunori, Yamatotakada-shi Nara-ken, JP;
Fujimoto, Yoshiji, Nara-shi Nara-ken, JP

(54) Einrichtung zur Sprachsynthese.

DE 689 15 353 T 2

EP-B-0 351 848
89 11 3343.1
Sharp Kabushiki Kaisha
u.Z.: Y 970 EP

Einrichtung zur Sprachsynthese

Die Erfindung betrifft eine Sprachsyntheseeinrichtung, die Wellensegmente, z.B. Tonwellensegmente und Quasisprachwellensegmente kompiliert, um eine Sprachwelle zu reproduzieren.

Es ist bekannt, daß von den verschiedenen Sprachwellen die Wellen von stimmhaften Tönen, z.B. von Vokalen, eine redundante Tonstruktur haben, bei der die gleiche Welle innerhalb einer Periode von 2 oder 3 ms bis zu 10 ms im wesentlichen mehrere Male bis zu ein Dutzend Male wiederholt wird. Herkömmlicherweise haben Sprachsynthesizer bisher ein Phonemsegmentkompilierverfahren unter Verwendung der oben erwähnten Tonstruktur verwendet, um eine synthetisierte Sprache zu erzeugen. Sprachsynthesizer dieser Art wiederholen und verbinden Tonwellensegmente oder Quasisprachwellensegmente für einen vorbestimmten Zeitraum, um eine Sprachwelle zu synthetisieren. Dies dient dazu, die Menge der Wellensegmentdaten für die Tonwellensegmente oder Quasisprachwellensegmente zu verringern, und es wird eine hohe Qualität der am Ende erreichten synthetisierten Sprache beibehalten.

Da jedoch ein herkömmlicher Sprachsynthesizer, der das Segmentkompilierverfahren, wie oben beschrieben, verwendet, eine Sprachwelle dadurch synthetisiert, daß er für eine vorbestimmte Zeitdauer einfach Tonwellensegmente oder Sprachwellensegmente, die auf den Tonwellensegmenten basieren, wiederholt oder verbindet, entstehen dort Verzerrungen, wo die Tonwellensegmente oder Quasisprachwellensegmente, wie oben beschrieben, verbunden werden.

Fig. 4 zeigt ein Beispiel für das Tonwellensegment, das bei der Sprachwellenformsynthese verwendet wird. Jeder Doppelkreis in Fig. 4 zeigt den abgetasteten Wert zu jeder Abtastzeit (nachstehend Abtastwert genannt); die durchgezogenen Linien, die von diesen Punkten senkrecht zur Zeitachse verlau-

fen, stellen die Abtastzeit dar; die gestrichelten Linien, die zwischen den Abtastpunkten senkrecht zur Zeitachse verlaufen, stellen die interpolierte Abtastzeit dar, zu der der Abtastwert interpoliert wird, um während der Wellenformsynthese den interpolierten Wert auszugeben. Das Tonwellensegment gemäß Fig. 4 kann je nach der Position, an der die Welle den Nullpunkt durchquert, einer der folgenden vier Wellentypen sein.

Insbesondere wird die Abtastzeitdauer Ts in zwei Phasen eingeteilt, wobei die erste mit P1 und die nächste mit P2 bezeichnet wird. Somit fällt beim Wellentyp (1) gemäß Fig. 4(a) der Nulldurchgangspunkt m für die interpolierte Wellenform des oberen oder (hier durchgängig austauschbar:) führenden Abtastwertes des Tonsegments in den Bereich P2 und der Nulldurchgangspunkt o für die interpolierte Wellenform des Endabtastwertes des Tonsegments in den Bereich P2. Beim Wellentyp (2) gemäß Fig. 4(b) fällt der Nulldurchgangspunkt für die interpolierte Wellenform des oberen oder führenden Abtastwertes des Tonsegments in den Bereich P1 und der Nulldurchgangspunkt für die interpolierte Wellenform des Endabtastwertes des Tonsegments in den Bereich P1. Beim Wellentyp (3) gemäß Fig. 4(c) fällt der Nulldurchgangspunkt für die interpolierte Wellenform des oberen Abtastwertes des Tonsegments in den Bereich P2 und der Nulldurchgangspunkt für die interpolierte Wellenform des Endabtastwertes des Tonsegments in den Bereich P1. Beim Wellentyp (4) gemäß Fig. 4(b) fällt der Nulldurchgangspunkt für die interpolierte Wellenform des oberen Abtastwertes des Tonsegments in den Bereich P1 und der Nulldurchgangspunkt für die interpolierte Wellenform des Endabtastwertes des Tonsegments in den Bereich P2. Wenn die Tonwellensegmente jedes der oben beschriebenen Wellentypen also einfach wiederholt und verbunden werden, wird die Tonperiode, wo die Segmente verbunden werden, um einen Betrag phasenverschoben, der der halben Abtastzeitdauer entspricht, was zur Verzerrung führt, was wiederum einen Unterschied zur ursprünglichen Welle bedeutet.

Das heißt, wenn beispielsweise gleiche Wellen des Typs (3) einfach verbunden werden, wird die Phase der resultierenden Welle um einen halben Abtastzyklus verzögert, wie in Fig.

5(b) dargestellt. Wenn ferner gleiche Wellen des Typs (4) einfach verbunden werden, wird die Phase der resultierenden Welle um einen halben Abtastzyklus nach vorn verschoben, wie in Fig. 5(c) dargestellt. In diesem Fall tritt beim Anstieg des Tonwellensegments Interferenz auf, und die Tonqualität der am Ende synthetisierten Sprache verschlechtert sich deutlich. Die Verschlechterung der Tonqualität ist besonders stark, wenn die Tonperiode kurz ist (d.h. wenn die Tonfrequenz hoch ist), wie das bei Frauenstimmen der Fall ist.

WO-A-85/04 747 beschreibt eine Text-Sprache-Umwandlung durch Steuern der Tonfrequenz, da die ursprüngliche Tonfrequenz der zu verbindenden Wellenformsegmente sich von der Tonfrequenz unterscheidet, die erforderlich ist, um die Sprachsynthese durchzuführen.

Um das oben erörterte Problem zu lösen, gibt es zwei Verfahren. Beim ersten Verfahren wird ein Tonwellensegment herausgetrennt, durch schnelle Fourier-Transformations-(FFT-)Analyse zeitweilig in eine Frequenzachsenwelle umgewandelt und mittels umgekehrter FFT nach Phasenkorrektur wieder in eine Zeitachsenwelle zurückverwandelt, so daß beide Enden des Tonwellensegments sich dem Wert Null nähern können. Beim anderen Verfahren wird mittels linearer prädiktiver Codierung (LPC) der einen Tonwelle, die herausgetrennt worden ist, eine Impulsantwortwelle erzeugt, und diese Impulsantwortwelle wird als Tonwellensegment verwendet. Bei den oben genannten Verfahren liegen jedoch die Enden des Tonwellensegments nicht nahe genug an dem Wert Null, und die Verzerrung bleibt somit im Tonwellensegment erhalten, was zu Veränderungen des Tons führt.

Es ist daher eine Aufgabe der Erfindung, eine Sprachsyntheseeinrichtung bereitzustellen, die durch ein einfaches Verfahren, bei dem die Wellensegmente verbunden werden, synthetische Sprache ohne Tonqualitätsstörungen erzeugt. Diese Aufgabe wird mit den Merkmalen der Ansprüche gelöst.

Eine Sprachsyntheseeinrichtung, wie beschrieben, kompiliert Wellensegmente, z.B. Tonwellensegmente der Sprache, um Sprache zu synthetisieren, und weist auf: einen Verbindungsty-

pspeicher zum Speichern eines Verbindungstyps, der den Verbindungszustand des Punktes beschreibt, an dem die Wellensegmente verbunden werden; und einen Wellensegmentverbinder, der beim Verbinden der Wellensegmente den Endabtastpunkt und den führenden Abtastpunkt der Wellensegmente mit einer herkömmlichen Abtastzeitdauer oder mit einer herkömmlichen Abtastzeitdauer, die entsprechend dem in Verbindungstypspeicher gespeicherten Verbindungstyp um lediglich 1/2 der Abtastzeitdauer zusammengedrückt oder gedehnt wird, verbindet.

Wenn also Sprachwellensegmente kompiliert werden, um Sprache zu synthetisieren, wird der im Verbindungstypspeicher gespeicherte Verbindungstyp abgefragt. Entsprechend dem abgefragten Verbindungstyp werden der End- und der führende Abtastpunkt der Wellensegmente verbunden mit einer herkömmlichen Abtastzeitdauer oder mit einer herkömmlichen Abtastzeitdauer, die um lediglich 1/2 der Abtastzeitdauer zusammengedrückt oder gedehnt wird, so daß die Wellensegmente übergangslos verbunden werden, um eine synthetische Sprachwelle bereitzustellen.

Die Erfindung wird nachstehend mit Bezug auf die Zeichnungen näher beschrieben. Dabei zeigen:

Fig. 1 ein Blockschaltbild einer bevorzugten Ausführungsform einer Sprachsyntheseeinrichtung gemäß der Erfindung;

Fig. 2 eine grafische Darstellung des Formats für die Speicherung der Tonwellensegmentdaten in einem Festwertspeicher (ROM);

Fig. 3 ein Ablaufdiagramm, das die Aufeinanderfolge des Ablaufs des Sprachsynthesevorgangs darstellt;

Fig. 4(a) bis Fig. 4(d) Zeichnungen zur Beschreibung der Wellentypen;

Fig. 5(a) bis Fig. 5(c) grafische Darstellungen zur Erläuterung der Wellentypen und ihrer Verbindungsverfahren;

Fig. 6(a) bis Fig. 6(d) grafische Darstellungen zur Erläuterung der Wellentypen gemäß einer alternativen erfindungsgemäßen Ausführungsform; und

Fig. 7(a) und 7(b) grafische Darstellungen zur Erläuterung der Wellentypen und ihrer Verbindungsverfahren gemäß einer alternativen erfindungsgemäßen Ausführungsform.

Die erste bevorzugte erfindungsgemäße Ausführungsform wird nachstehend mit Bezug auf Fig. 1 beschrieben, die ein Blockschaltbild einer erfindungsgemäßen Sprachsyntheseeinrichtung darstellt.

Bezugszeichen 1 bezeichnet einen Steuerungs-ROM (Festwertspeicher), der ein Steuerprogramm speichert, das von der CPU (zentrale Verarbeitungseinheit) 5 zur Sprachsynthese verwendet wird; Bezugszeichen 2 bezeichnet einen RAM (Direktzugriffsspeicher), der als Arbeitsspeicher während der Sprachsynthese verwendet wird; Bezugszeichen 3 bezeichnet einen Daten-ROM, der verwendet wird, um Sprachcodierungsdaten zu speichern; Bezugszeichen 4 bezeichnet eine E/A-Schnittstelle, durch die zu Beginn der Sprachsynthese und anderer Vorgänge Eingangs/Ausgangssignale laufen; Bezugszeichen 6 bezeichnet einen D/A-Wandler, der zur Digital-Analog-Wandlung von Sprachwellendaten, die unter der Steuerung der CPU synthetisiert werden, verwendet wird; und Bezugszeichen 7 bezeichnet einen Verstärker, der eine analoge Eingangssprachwelle verstärkt und sie an einen Lautsprecher 8 übergibt.

Der Steuerungs-ROM 1, der RAM 2, der Daten-ROM 3, die E/A-Schnittstelle 4, die CPU 5 und der D/A-Wandler 6, die alle in der Sprachsyntheseeinrichtung mit dem oben beschriebenen Aufbau verwendet werden, können auf einem einzigen Chip integriert sein. Es ist auch möglich, einen externen Daten-ROM 9 zum Speichern der Sprachcodierungsdaten als Systemerweiterung zu verwenden.

Wenn ein Startsignal, das erforderlich ist, um die Sprachsynthese auszulösen, in die Sprachsyntheseeinrichtung mit dem oben beschriebenen Aufbau über die E/A-Schnittstelle 4 aus einer externen Quelle eingegeben wird, beginnt die CPU 5 den Sprachsynthesevorgang, der auf dem Steuerungsprogramm beruht, das im Steuerungs-ROM 1 gespeichert ist. Dabei werden von der CPU 5 Sprachsynthesewellendaten erzeugt, die auf den Sprachcodierungsdaten basieren, die im Daten-ROM 3 gespeichert sind. Die erzeugten Sprachsynthesewellendaten werden vom D/A-Wandler 6 in ein Analogsignal umgewandelt, dann vom Verstärker 7 verstärkt und schließlich als synthetisierte Sprache vom Lautsprecher 8 ausgegeben.

Wie weiter unten beschrieben, erzeugt die erfindungsgemäße Sprachsyntheseeinrichtung synthetisierte Sprache, die frei von Störungen des Tonwellenanstiegs ist, indem Wellensegmente, z.B. Tonwellensegmente oder Quasisprachwellensegmente, verbunden werden, um die synthetisierte Sprache zu erzeugen.

Wenn bei einem ersten Verfahren gemäß Fig. 5(a) der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den Endabtastwert des vorangegangenen Tonwellensegments und der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den oberen Abtastwert des folgenden Tonwellensegments beide im Bereich P2 liegen, wenn die Wellen aufgrund der Verbindung gleicher Wellen des Typs (1) oder ungleicher Wellen des Typs (1) und des Typs (3), wie in Fig. 4(a) und 4(c) dargestellt, und wenn der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den Endabtastwert des vorangegangenen Tonwellensegments und der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den oberen Abtastwert des folgenden Tonwellensegments beide im Bereich P1 liegen, wenn die Wellen aufgrund der Verbindung gleicher Wellen des Wellentyps (2) oder ungleicher Wellen des Wellentyps (2) und des Wellentyps (4) verbunden werden, werden der Endabtastwert und der obere Abtastwert der Tonwellensegmente am herkömmlichen Abtastpunkt ausgegeben, und die Tonwellensegmente werden verbunden. Danach werden die interpolierten Werte zwischen dem Endabtastwert und dem oberen Abtastwert (dargestellt durch ein Dreieck mit durchgezogenen Linien) in einem Punkt berechnet, der gleich 1/2 Abtastintervall Ts ist und so ausgegeben, daß die beiden Tonwellensegmente übergangslos verbunden werden können. Im folgenden wird die Verbindung von solchen Tonwellensegmenten, wie eben beschrieben, als Verbindungstyp 0a bezeichnet.

Wenn, wie in Fig. 5(b) dargestellt, der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den Endabtastwert des vorangegangenen Tonwellensegments im Bereich P1 liegt und der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den oberen Abtastwert des folgenden Tonwellensegments im Bereich P2 liegt, wenn die Wellen aufgrund der Verbindung ungleicher Wellen des Typs (2) und des Typs (1)

oder Wellen des Typs (2) und des Typs (3) verbunden werden, werden die Wellensegmente nicht am herkömmlichen Abtastpunkt verbunden; das herkömmliche Abtastintervall zwischen dem Endabtastpunkt und dem oberen Abtastpunkt wird um 1/2 zusammengedrückt und wird dann ausgegeben, um die Tonwellensegmente zu verbinden. Im folgenden wird die Verbindung von solchen Tonwellensegmenten, wie eben beschrieben, als Verbindungstyp 1a bezeichnet.

Wenn, wie in Fig. 5(c) dargestellt, der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den Endabtastwert des vorangegangenen Tonwellensegments im Bereich P2 liegt und der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den oberen Abtastwert des folgenden Tonwellensegments im Bereich P1 liegt, wenn die Wellen aufgrund der Verbindung von ungleichen Wellen des Typs (1) und des Typs (2) oder von Wellen des Typs (1) und des Typs (4) verbunden werden, werden die Wellensegmente im herkömmlichen Abtastpunkt nicht verbunden; das herkömmliche Abtastintervall zwischen dem Endabtastwert und dem oberen Abtastwert wird um 1/2 gedehnt und dann ausgegeben, um die Tonwellensegmente zu verbinden. Der Zeitraum zwischen dem Endabtastwert und dem oberen Abtastwert der Tonwellensegmente wird wie folgt interpoliert.

Wenn wir insbesondere annehmen, daß der Endabtastwert des vorangegangenen Tonwellensegments $|x1|$ beträgt und der obere Abtastwert des folgenden Tonwellensegments $|x2|$ beträgt, wenn $|x1|>|x2|$, wird der interpolierte Wert $x1/2$ nach dem Endabtastwert $|x1|$ (insbesondere dem höheren Spitzenwert) berechnet und wird dann in Intervallen von $Ts/2$ ausgegeben. Als nächstes wird der Zeitraum zwischen diesem interpolierten Wert $x1/2$ und dem oberen Abtastwert $|x2|$ (insbesondere dem unteren Spitzenwert) interpoliert und dann ausgegeben. Im folgenden wird die Verbindung von solchen Tonwellensegmenten, wie eben beschrieben, als Verbindungstyp 2-(a) bezeichnet. Wenn ferner $|x1|<|x2|$, wird der interpolierte Wert $x2/2$ des vorausgegangenen oberen Abtastwertes $|x2|$ berechnet und dann in Intervallen von $Ts/2$ ausgegeben. Als nächstes wird der Zeitraum zwischen diesem interpolierten Wert $x2/2$ und dem oberen Abtastwert $|x1|$ (insbesondere dem unteren Spitzenwert)

interpoliert und dann ausgegeben. Im folgenden wird die Verbindung von solchen Tonwellensegmenten, wie eben beschrieben, als Verbindungstyp 2-(b) beschrieben.

Bei dem zweiten Verfahren wird das Abtasten in einem Zyklus durchgeführt, der zweimal (das Zweifache der Frequenz) so groß ist, wie im Nyquist-Theorem definiert. Unabhängig davon, ob die Abtastung in einem geradzahligen Abtastpunkt oder einem ungeradzahligen Abtastpunkt stattfindet, werden die Abtastdaten, die zur Sprachsynthese verwendet werden, im Standardzyklus des Nyquist-Theorems vom Abtastpunkt an, der dem Anstieg des Tonsegments am nächsten ist, erneut abgetastet. Diese Welle ist in Fig. 6(a) bis 6(b) dargestellt. Hier sind die geradzahligen Abtastpunkte die Abtastpunkte (dargestellt durch eine durchgezogene Linie in Fig. 6), die im Nyquist-Theorem-Zyklus auftreten, und die ungeradzahligen Abtastpunkte (dargestellt durch eine gestrichelte Linie in Fig. 6) sind die Abtastpunkte, die zwischen den geradzahligen Abtastpunkten auftreten. In diesem Fall sind die Abtastdaten, die in den Abtastpunkten ermittelt werden, die durch einen Doppelkreis dargestellt werden, die Abtastpunkte (im folgenden als Zielabtastpunkte bezeichnet), die das Ziel der Sprachsynthese sind. Diese Segmente können entweder Wellentyp (1) oder Wellentyp (2) sein.

Wenn, wie in Fig. 7(a) dargestellt, der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den Endabtastwert, der das Ziel der Sprachsynthese für das vorangegangene Tonwellensegment ist (im folgenden Endzielabtastwert genannt) und der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den führenden Abtastwert des folgenden Tonwellensegments beide aufgrund der Verbindung von gleichen Wellen des Typs (5) oder ungleichen Wellen des Typs (5) und des Typs (6) im Bereich P2 liegen, werden der Endzielspitzenwert, der das Ziel der Sprachsynthese ist, und der führende Zielabtastwert an dem Abtastpunkt ausgegeben, der das Ziel der Sprachsynthese ist, um die Tonwellensegmente zu verbinden. Danach wird am halben Punkt der Zielabtastzeitdauer der Endabtastwert q des vorangegangenen Tonwellensegments als der interpolierte Wert ausgegeben, so daß die beiden Tonwellensegmente über-

gangslos verbunden werden können. Im folgenden wird die Verbindung von solchen Tonwellensegmenten als Verbindungstyp 0b bezeichnet.

Wenn, wie in Fig. 7(b) dargestellt, der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den Endzielabtastwert des vorangegangenen Tonwellensegments im Bereich P1 liegt und der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den führenden Zielabtastwert des folgenden Tonwellensegments im Bereich P2 liegt, werden aufgrund der Verbindung von gleichen Wellen des Typs (6) oder ungleichen Wellen des Typs (6) und des Typs (5) die Tonwellensegmente nicht an dem Abtastpunkt verbunden, der das Ziel der Sprachsynthese ist; die Zeitdauer zwischen dem Endzielabtastwert und dem führenden Zielabtastwert der Tonwellensegmente wird um 1/2 zusammengedrückt und dann ausgegeben, um die Tonwellensegmente zu verbinden. Im folgenden wird die Verbindung von solchen Tonwellensegmenten als Verbindungstyp 1b bezeichnet.

Fig. 2 zeigt ein Beispiel des Datenformats, wenn z.B. die Tonwellensegmente analysiert und die resultierenden Tonwellensegmentdaten im ROM 3 gespeichert sind (siehe Fig. 1). Das dargestellte Datenformat besteht aus Codierungsdaten aus mehreren Tonwellensegmenten, wobei die einzelnen Codierungsdaten für jedes Tonwellensegment Interpolationsdaten und Sprachdaten aufweisen. Die Interpolationsdaten bestehen aus Endsegmentdaten 11, die anzeigen, ob das Tonwellensegment das letzte Tonwellensegment ist oder nicht, Codierungsverfahrensdaten 12, die das Verfahren anzeigen, das verwendet wird, um die Abtastdaten des Tonwellensegments zu codieren, Wiederholungsanzahldaten 13, die anzeigen, wie oft das Tonwellensegment wiederholt wurde, Verbindungstypdaten 14, wie in Fig. 5 und Fig. 7 dargestellt, zur Verwendung, wenn das gleiche Tonwellensegment wiederholt wird, und Verbindungstypdaten 15 (im folgenden als Folgetonwellensegmentverbindungstyp bezeichnet) zur Verwendung, wenn das gegebene Tonwellensegment mit dem nächsten angrenzenden Tonwellensegment verbunden ist. Die Sprachdaten weisen Abtastwertanzahldaten 16, die die Anzahl der codierten Daten angeben, die im Tonwellensegment enthalten sind, und eine Serie von mehreren codierten Daten 17 bis 19

für jeden Abtastpunkt, der in der Sprachsynthese verwendet wird, auf. Diese codierten Daten werden als eine Bitfolge entsprechend dem Codierungsverfahren (z.B. Pulscodemodulation (PCM) oder adaptive Differenzpulscodemodulation (ADPCM)) gespeichert, die in den Codierungsverfahrensdaten 12 für die Interpolationsdaten gespeichert ist.

Mit Bezug auf das Ablaufdiagramm gemäß Fig. 3 wird nachstehend der Sprachsynthesevorgang genauer beschrieben, bei dem die Tonwellensegmente, die Wellensegmente sind, verbunden werden und Sprache mittels der Verfahren 1 und 2, die oben beschrieben worden sind, synthetisiert wird.

In Schritt S1 wird ein Byte der Interpolationsdaten aus den Tonwellensegmentdaten gelesen, die im Daten-ROM 3 entsprechend dem Format gemäß Fig. 2 gespeichert sind, und das Byte wird in die Endsegmentdaten 11, die Codierungsverfahrensdaten 12, die Wiederholungsanzahldaten 13, die Verbindungstypdaten 14 und den Folgetonwellensegmentverbindungstyp 15 eingeteilt. Auf der Grundlage der gewonnenen Informationen werden jeweils das Endsegmentdaten-Flag, das Codierungsverfahren-Flag, der Wiederholungszähler, der Wiederholungsverbindungstyp und der Folgetonwellensegmentverbindungstyp im RAM 2 gesetzt. Der RAM 2 hat einen Bereich zum Speichern des Wiederholungsverbindungstyps zur Wellensegmentverbindung und eines Tonwellensegmentverbindungstyps zur Wellensegmentverbindung, und der Wiederholungsverbindungstyp der vorangegangenen Tonwellensegmentdaten und der Folgetonwellensegmentverbindungstyp werden beide dort gesetzt.

In Schritt S2 werden Abtastwertanzahldaten 16, die die Anzahl der codierten Daten eines Tonwellensegments angeben, aus den Daten-ROM 3 gelesen, und diese Anzahl wird im RAM 2 als Abtastwertanzahl gesetzt.

In Schritt S3 wird das erste codierte Datenelement aus den Daten-ROM 3 gelesen.

In Schritt S4 wird das erste codierte Datenelement entsprechend dem Codierungsverfahren decodiert, die im Codierungsverfahren-Flag des RAM 2 gesetzt worden ist, und der obere Abtastwert des Tonwellensegments wird berechnet. Der interpolierte Wert der Zeitdauer zwischen diesem oberen Abtast-

wert und dem folgenden Abtastwert (auf der Grundlage des zweiten codierten Datenelements) wird dann berechnet. Als nächstes wird der Interpolationsvorgang, der zum Verbinden mit dem vorangegangenen Tonwellensegment erforderlich ist, entsprechend dem Folgetonwellensegmentverbindungstyp der vorangegangenen Tonwellensegmentdaten ausgeführt, die im Wiederholungsverbindungstyp für Tonwellensegmente im RAM 2 gesetzt worden sind. Ferner wird der Zeitablauf für die Ausgabe des berechneten oberen Abtastwertes an den D/A-Wandler 6 berechnet (wenn der Verbindungstyp 0a und 0b ist, wird der normale Zeitablauf ausgegeben; wenn der Verbindungstyp 1a und 1b ist, wird der Zeitablauf eines Abtastzyklus ausgegeben, der um 1/2 nach vorn verschoben ist; wenn der Verbindungstyp 2a und 2b ist, wird der Zeitablauf eines Abtastzyklus ausgegeben, der um 1/2 verzögert ist).

In Schritt S5 werden der obere Abtastwert, der in Schritt S4 berechnet worden ist, und der Ausgabezeitablauf der vorangegangenen und der folgenden interpolierten Werte, die in Schritt S4 berechnet worden sind, an den D/A-Wandler 6 übergegeben.

Das heißt, die Interpolation wird entsprechend den vier Verbindungstypen, die in Fig. 5 dargestellt sind, durchgeführt und zwar unabhängig davon, ob die Zeitdauer zwischen dem Endabtastwert des vorangegangenen Tonwellensegments und dem oberen Abtastwert des gegenwärtigen Tonwellensegments um 1/2 Abtastzyklus gedehnt oder zusammengedrückt worden sind, und danach findet die D/A-Wandlung statt.

In Schritt S6 werden die nächsten codierten Daten (die zweiten codierten Daten) aus dem Daten-ROM 3 gelesen.

In Schritt S7 werden die nächsten codierten Daten entsprechend dem Codierungsverfahren decodiert, und der nächste Abtastwert wird berechnet. Danach wird der interpolierte Wert der Zeitdauer bis zum nächsten Abtastwert berechnet. Der berechnete Abtastwert und der interpolierte Wert werden mit dem normalen Zeitablauf (insbesondere am normalen Abtastpunkt) übergeben.

In Schritt S8 wird der Abtastwertzähler um 1 erhöht, und es wird aufgrund dieses Wertes festgestellt, ob die

Verarbeitung der codierten Daten des augenblicklichen Tonwellensegments beendet worden ist oder nicht. Wenn festgestellt wird, daß die gesamte Verarbeitung beendet worden ist, geht der Ablauf weiter mit dem Schritt S9; wenn nicht, erfolgt eine Rückkehr nach Schritt S6; und in beiden Fällen wird die Verarbeitung der nächsten codierten Daten ausgeführt.

In Schritt S9 wird der Wiederholungsverbindungstyp der vorangegangenen Tonsegmentdaten, der im Wiederholungsverbindungstyp für Tonwellensegmente in RAM 2 gesetzt worden ist, zurückgesetzt auf den Wiederholungsverbindungstyp der gegenwärtigen Tonwellensegmentdaten, der im Wiederholungsverbindungstyp in RAM 2 gesetzt worden ist.

In Schritt S10 wird der Wiederholungszähler in RAM 2 um 1 verringert, und es wird auf der Grundlage dieses Wertes festgestellt, ob alle Wiederholungen des gegenwärtigen Tonwellensegments beendet sind oder nicht. Wenn Beendigung festgestellt wird, geht der Ablauf weiter mit dem Schritt S11; wenn nicht, erfolgt eine Rückkehr nach Schritt S3, die ersten codierten Daten des gegenwärtigen Tonwellensegments werden wiederum eingegeben, und eine erneute Verarbeitung wird ausgeführt.

In Schritt S11 wird der nächste Tonwellensegmentverbindungstyp der vorangegangenen Tonwellensegmentdaten, der im nächsten Tonwellensegmentverbindungstyp für Tonwellensegmente in RAM 2 gesetzt worden ist, zurückgesetzt auf den nächsten Tonwellensegmentverbindungstyp der gegenwärtigen Tonwellensegmentdaten, der im Folgetonwellensegmentverbindungstyp von RAM 2 gesetzt worden sind.

Im Schritt S12 wird das Endsegmentdaten-Flag in RAM 2 abgefragt, um festzustellen, ob das gegenwärtige Tonwellensegment das Endsegment ist. Wenn ja, wird der Sprachsynthesevorgang beendet; wenn nein, erfolgt eine Rückkehr nach Schritt S1, die nächsten Tonwellensegmentdaten werden gelesen, und die Verarbeitung der nächsten Tonwellensegmentdaten beginnt.

Somit werden die Wellensegmentverbindungstypen anhand der Kombination der Verbindungen von Tonwellensegmenten verschiedener Wellentypen kategorisiert. Auf der Grundlage des Verbindungstyps kann die Zeitdauer zwischen dem Endabtastpunkt

und dem führenden Abtastpunkt der verbundenen Tonwellensegmente um 1/2 der normalen Abtastzeitdauer zusammengedrückt oder gedehnt werden, oder es kann die normale Abtastzeitdauer verwendet werden, um die Wellensegmente zu verbinden. Somit können die Tonwellensegmente durch einen einfachen Vorgang übergangslos verbunden werden, ohne daß eine Phasenverschiebung bei der Verbindung der Tonwellensegmente erzeugt wird. Das heißt, bei einer Sprachsyntheseeinrichtung gemäß der Erfindung tritt beim Anstieg des Tonwellensegments keine Verzerrung auf, und es wird keine Tonqualitätsverschlechterung hervorgerufen.

Bei der bevorzugten Ausführungsform, wie oben beschrieben, wird ein Tonwellensegment als Wellensegment verwendet, jedoch ist die Erfindung nicht darauf beschränkt, und ein Sprachwellensegment, das einem Tonwellensegment entspricht, kann ebenfalls verwendet werden.

Wie aus der vorangegangenen Beschreibung der Erfindung bekannt ist, treten bei der Verbindung von Wellensegmenten bei der synthetischen Sprache, die durch die erfindungsgemäße Sprachsyntheseeinrichtung erzeugt wird, keine Phasenverschiebungen auf. Dieser Vorteil beruht darauf, daß die Sprachsyntheseeinrichtung mit dem Wellensegmentverbinder ausgestattet ist, der einen Verbindungstyp speichert, der den Typ der Verbindung zwischen den Wellensegmenten in der Sprache in einem Verbindungstypspeicher speichert. Wenn ferner die Wellensegmente verbunden werden, um Sprache zu synthetisieren, werden der Endabtastpunkt und der führende Abtastpunkt der Wellensegmente entsprechend dem Verbindungstyp, der im Verbindungstypspeicher gespeichert ist, um eine normale Abtastzeitdauer oder um eine Abtastzeitdauer, die um 1/2 der Zeitdauer zusammengedrückt oder gedehnt ist, miteinander verbunden.

Dadurch kann der Zeitraum zwischen Tonwellensegmenten interpoliert werden, und die Segmente können durch einen einfachen Vorgang übergangslos verbunden werden. Somit kann durch die Erfindung Sprachsynthese, die frei von Verzerrungen im Anstieg der verbundenen Wellensegmente ist und keine Verschlechterung der Tonqualität aufweist, erreicht werden.

Obwohl die Erfindung im Zusammenhang mit den bevorzugten Ausführungsformen mit Bezug auf die beigefügten Zeichnungen vollständig beschrieben worden ist, beachte man, daß für den Fachmann verschiedene Änderungen und Modifikationen offensichtlich sind. Solche Veränderungen und Modifikationen gelten als in den Umfang der Erfindung eingeschlossen, der in den beigefügten Patentansprüchen definiert ist.

EP-B-0 351 848
89 11 3343.1
Sharp Kabushiki Kaisha
u.Z.: Y 970 EP

Patentansprüche

1. Sprachsyntheseeinrichtung zum Verbinden von Wellensegmenten, um eine synthetisierte Sprache zu erzeugen, die frei von Verzerrungen im Tonwellenanstieg ist, mit:

a) einem Verbindungstypspeicher zum Speichern mehrerer bevorzugter Verbindungstypen für Wellensegmente, wobei die Verbindungstypen jeweils eine Verbindung einer interpolierten Wellenform für einen Endabtastwert eines vorangegangenen Wellensegments eines bestimmten Typs mit einer interpolierten Wellenform für einen führenden Abtastwert eines folgenden Wellensegments eines bestimmten Typs darstellt, wobei jeder der bevorzugten Verbindungstypen eine bevorzugte Abtastzeitdauer zur Verwendung während der Verbindung der Wellensegmente festlegt; und

b) einem Wellensegmentverbinder zum Festlegen bestimmter Typen von Wellensegmenten zum Vergleich mit den mehreren bevorzugten Verbindungstypen durch Interpolieren eines Zeitachsennulldurchgangspunkts für die interpolierte Wellenform für den Endabtastwert des vorangegangenen Wellensegments und eines Zeitachsennulldurchgangspunkts für die interpolierte Wellenform für den führenden Abtastwert des folgenden Wellensegments, wobei der Wellensegmentverbinder Verbindung der Wellensegmente unter Verwendung einer der bevorzugten Abtastzeitdauern herstellt.

2. Sprachsyntheseeinrichtung nach Anspruch 1, wobei die bevorzugte Abtastzeitdauer eine Abtastzeitdauer aufweist, die aus einer Gruppe ausgewählt wird, die aus einer vorbestimmten Abtastzeitdauer, einem Zweifachen einer vorbestimmten Abtastzeitdauer und einer Hälfte einer vorbestimmten Abtastzeitdauer besteht.

3. Sprachsyntheseeinrichtung nach Anspruch 1 oder 2, wobei die besagten mehreren bevorzugten Verbindungstypen aufweisen:

a) einen ersten Verbindungstyp, bei dem sowohl der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den führenden Abtastwert des folgenden Wellensegments als auch der Zeitachsennulldurchgangspunkt des interpolierten Wellensegments für den Endabtastwert des vorangegangenen Wellensegments innerhalb einer zweiten Hälfte einer vorbestimmten Abtastzeitdauer liegen;

b) einen zweiten Verbindungstyp, bei dem sowohl der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den führenden Abtastwert des folgenden Wellensegments als auch der Zeitachsennulldurchgangspunkt des interpolierten Wellensegments für den Endabtastwert des vorangegangenen Wellensegments in einer ersten Hälfte einer vorbestimmten Abtastzeitdauer liegen;

c) einen dritten Verbindungstyp, bei dem der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den führenden Abtastwert des folgenden Wellensegments in einer zweiten Hälfte einer vorbestimmten Abtastzeitdauer liegt und der Zeitachsennulldurchgangspunkt des interpolierten Wellenformsegments für den Endabtastwert des vorangegangenen Wellensegments in einer ersten Hälfte einer vorbestimmten Abtastzeitdauer liegt;

d) einen vierten Verbindungstyp, bei dem der Zeitachsennulldurchgangspunkt der interpolierten Wellenform für den führenden Abtastwert des folgenden Wellensegments in einer ersten Hälfte einer vorbestimmten Abtastzeitdauer liegt und der Zeitachsennulldurchgangspunkt des interpolierten Wellensegments für den Endabtastwert des vorangegangenen Wellensegments in einer zweiten Hälfte einer vorbestimmten Abtastzeitdauer liegt.

4. Sprachsyntheseeinrichtung nach Anspruch 1, 2 oder 3, bei dem die Wellensegmente Sprachtonsegmente aufweisen.

5. Sprachsyntheseeinrichtung nach einem der Ansprüche 1 bis 4, wobei die Wellensegmente Sprachwellensegmente aufweisen.

6. Sprachsyntheseeinrichtung nach Anspruch 5, wobei die Sprachwellensegmente Quasisprachwellensegmente aufweisen.

7. Sprachsyntheseeinrichtung nach einem der Ansprüche 1 bis 6, wobei eine Festwertspeichereinrichtung ein Steuerungsprogramm zur Verwendung durch eine zentrale Verarbeitungseinheit für Sprachsynthese speichert, eine Zufallsachsenspeichereinrichtung als Arbeitsspeicher während der Sprachsynthese verwendet wird, eine Festwertdatenspeichereinrichtung verwendet wird, um Sprachcodierungsdaten zu speichern, eine Eingabe/Ausgabe-Schnittstelle vorhanden ist, über die zu Beginn der Sprachsynthese und während anderer Vorgänge Eingangs/Ausgangssignale laufen, ein Digital-Analog-Wandler zum Umwandeln von Sprachwellendaten verwendet wird, die unter der Steuerung der zentralen Verarbeitungseinheit synthetisiert worden sind, und wobei ein Verstärker eine analoge Eingangssprachwelle verstärkt und an einen Lautsprecher übergibt.

8. Verfahren zum übergangslosen Verbinden von Wellensegmenten, um synthetische Sprache zu erzeugen, die frei von Verzerrungen im Tonwellenanstieg ist, mit den Schritten:

a) Identifizieren eines Zeitachsennulldurchgangspunkts für eine interpolierte Wellenform eines Endabtastwertes eines vorangegangenen Wellensegments;

b) Festlegen eines Zeitachsennulldurchgangspunkts für eine interpolierte Wellenform eines führenden Abtastwertes eines folgenden Wellensegments;

c) Vergleichen der Zeitachsennulldurchgangspunkte des vorangegangenen Wellensegments und des folgenden Wellensegments mit einem Verbindungstypspeicher, um einen bevorzugten Wellensegmentsverbindungstyp auszuwählen;

d) Auswählen eines bevorzugten Wellensegmentsverbindungstyps und einer bevorzugten Abtastzeitdauer; und

e) Verbinden des vorangegangenen Wellensegments mit dem folgenden Wellensegment unter Verwendung des ausgewählten bevorzugten Wellensegmentverbindungstyps und der ausgewählten bevorzugten Abtastzeitdauer, um eine synthetisierte Sprache herzustellen, die unabhängig ist von Verzerrungen im Tonwellenanstieg.

9. Verfahren nach Anspruch 8, wobei der Schritt des Auswählens eines bevorzugten Wellensegmentverbindungstyps und einer bevorzugten Abtastzeitdauer die Schritte aufweist:

a) Kategorisieren der kombinierten Zeitachsennulldurchgangspunkte von jeder der interpolierten Wellenformen für das vorangegangene Wellensegment und das folgende Wellensegment zum Anpassen an die ähnlichste Speicherwellenform, die im Wellensegmentverbindungstypspeicher gespeichert ist; und

b) Interpolieren der bevorzugten Abtastzeitdauer entsprechend dem bevorzugten Wellensegmentverbindungstyp aus den Abtastzeitdauerauswahldaten, die vorbestimmte Abtastzeitdaten, ein Zweifaches der vorbestimmten Abtastzeitdaten und eine Hälfte der vorbestimmten Abtastzeitdaten aufweisen.

EP-B-0 351 848
89 11 3343.1
Sharp Kabushiki Kaisha
u.Z.: Y 970 EP

Fig. 1

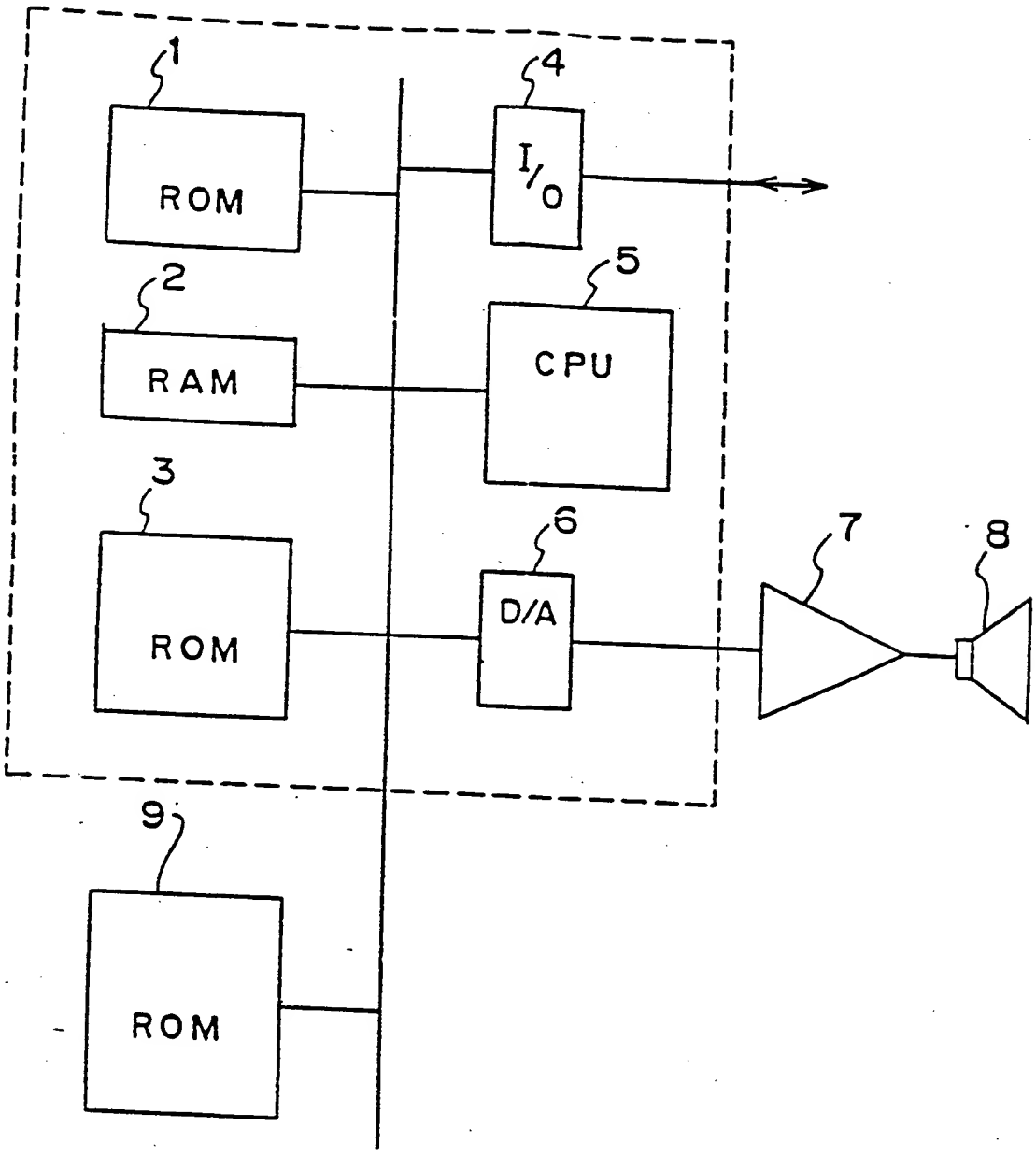


Fig. 2

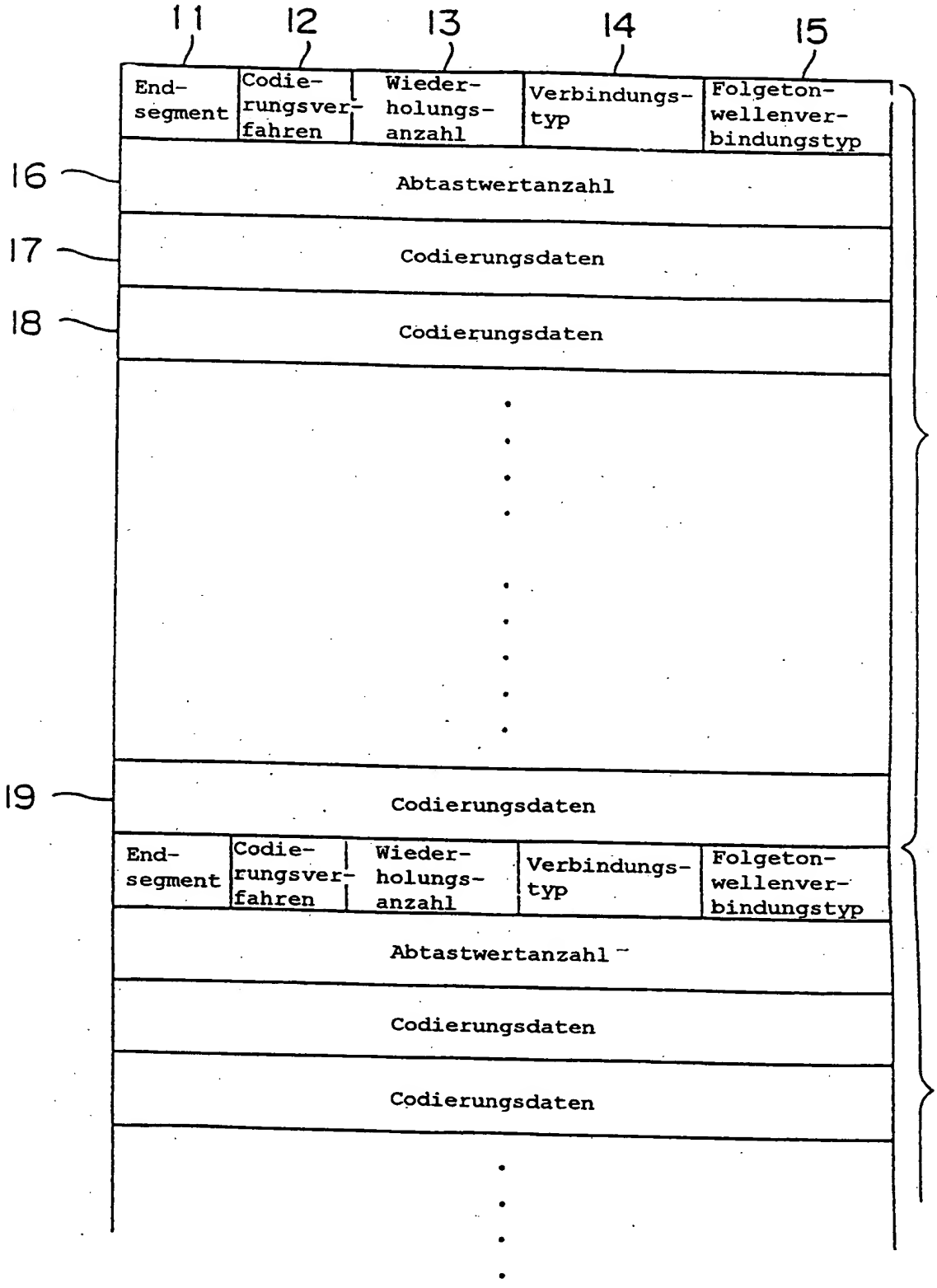

11
12
13
14
15
End-segment
Codierungsverfahren
Wiederholungsanzahl
Verbindungstyp
Folgetonwellenverbindungstyp
16
Abtastwertanzahl
17
Codierungsdaten
18
Codierungsdaten
19
Codierungsdaten
End-segment
Codierungsverfahren
Wiederholungsanzahl
Verbindungstyp
Folgetonwellenverbindungstyp
Abtastwertanzahl
Codierungsdaten
Codierungsdaten

Fig. 3

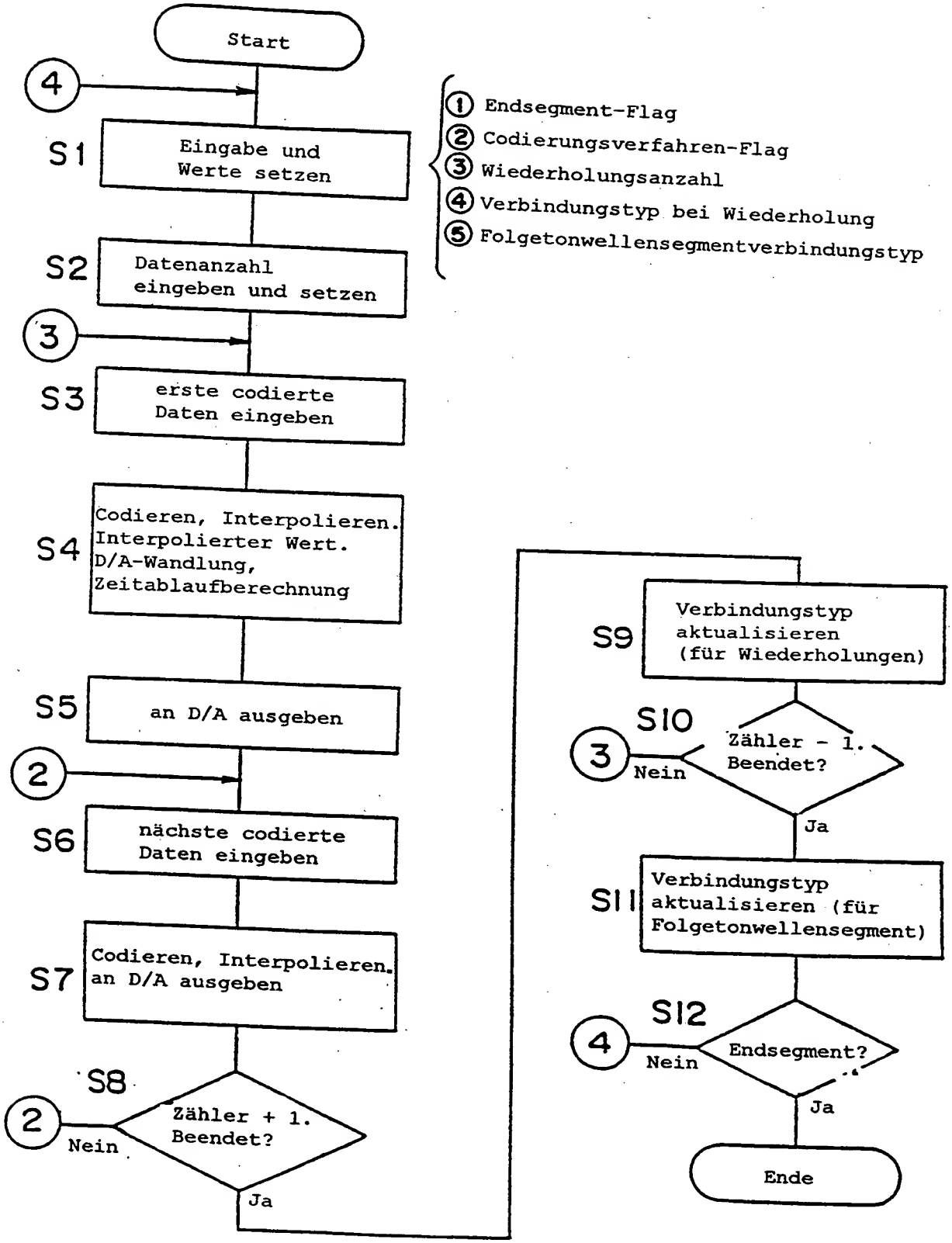

Start
4
S1
Eingabe und Werte setzen
S2
Datenanzahl eingeben und setzen
3
S3
erste codierte Daten eingeben
S4
Codieren, Interpolieren. Interpolierter Wert. D/A-Wandlung, Zeitablaufberechnung
S5
an D/A ausgeben
2
S6
nächste codierte Daten eingeben
S7
Codieren, Interpolieren. an D/A ausgeben
S8
2
Nein
Zähler + 1. Beendet?
Ja
S9
Verbindungstyp aktualisieren (für Wiederholungen)
S10
3
Nein
Zähler - 1. Beendet?
Ja
S11
Verbindungstyp aktualisieren (für Folgetonwellensegment)
S12
4
Nein
Endsegment?
Ja
Ende
① Endsegment-Flag
② Codierungsverfahren-Flag
③ Wiederholungsanzahl
④ Verbindungstyp bei Wiederholung
⑤ Folgetonwellensegmentverbindungstyp

Fig. 4

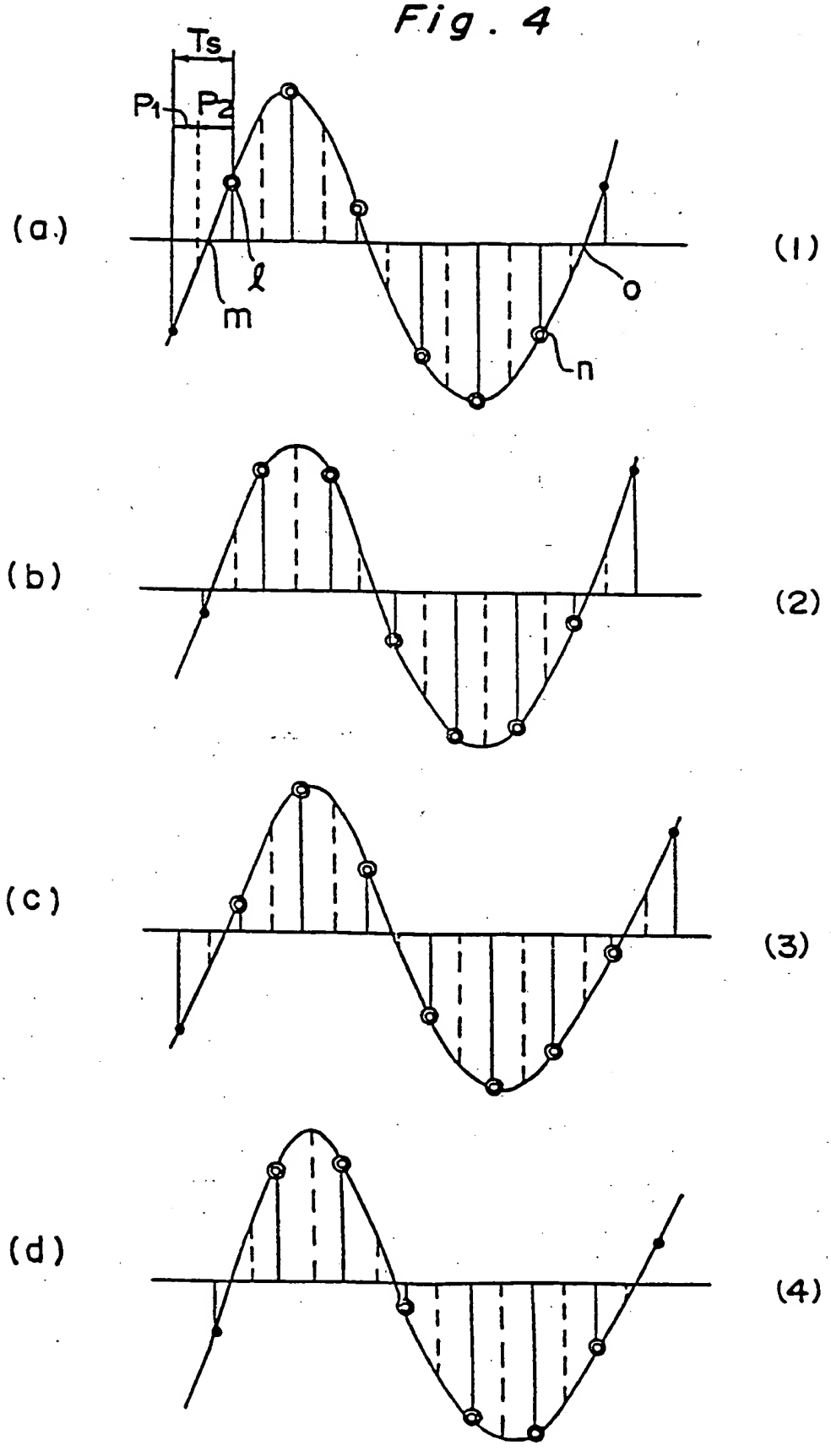
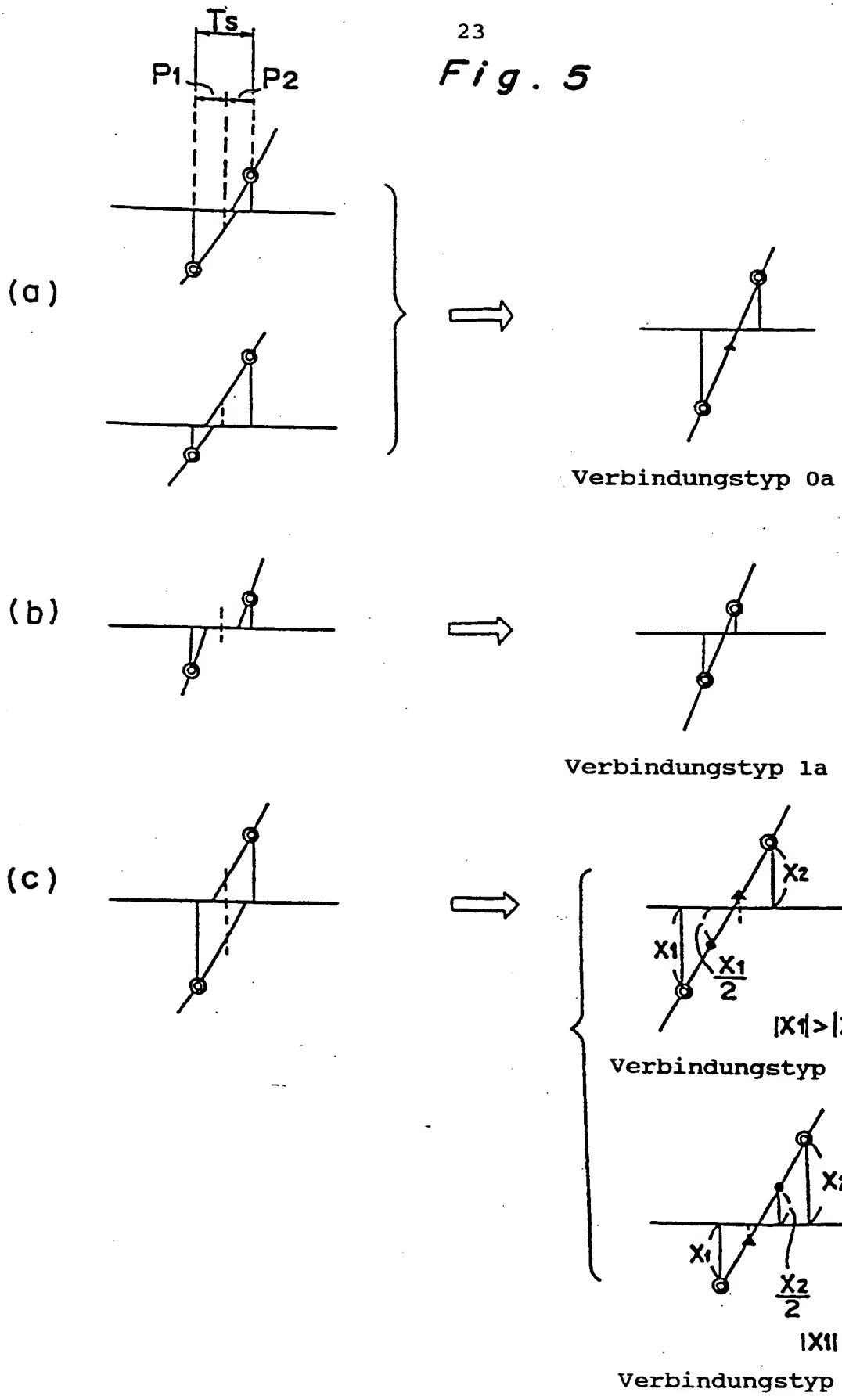


Fig. 5



Verbindungstyp 0a

Verbindungstyp 1a

Verbindungstyp 2-(a)

Verbindungstyp 2-(b)

Fig. 6

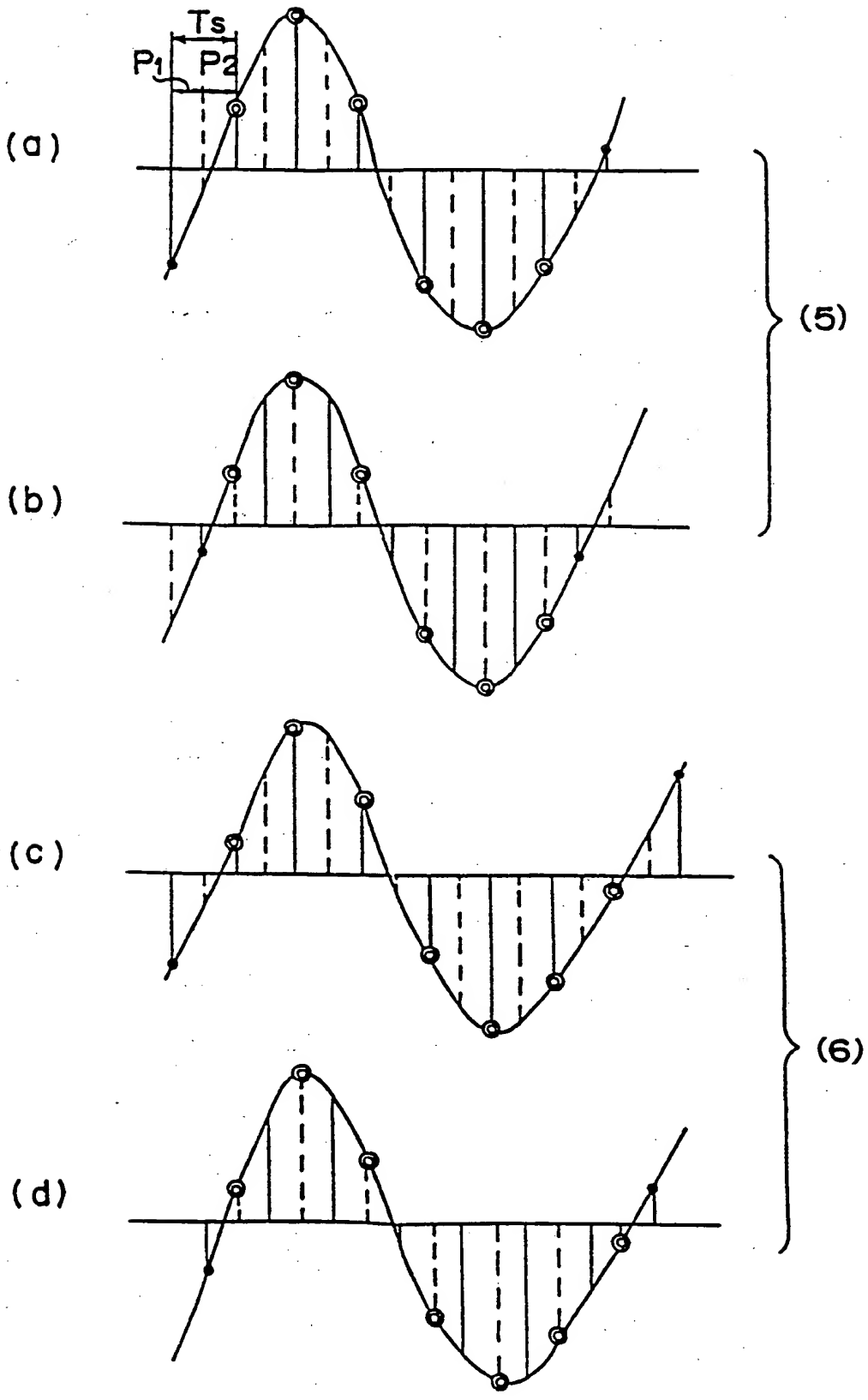
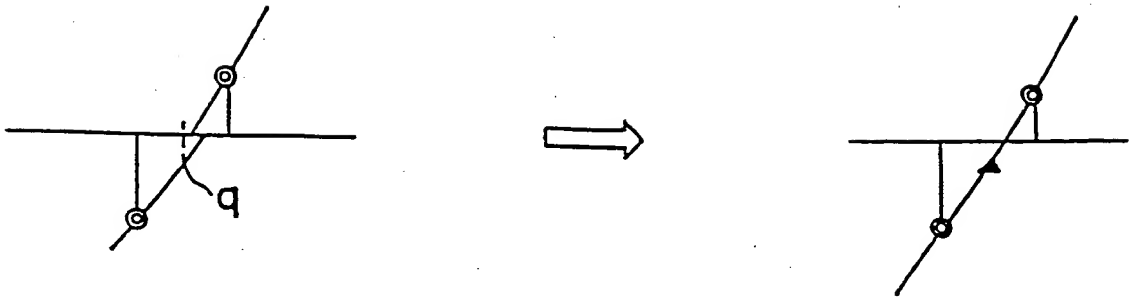
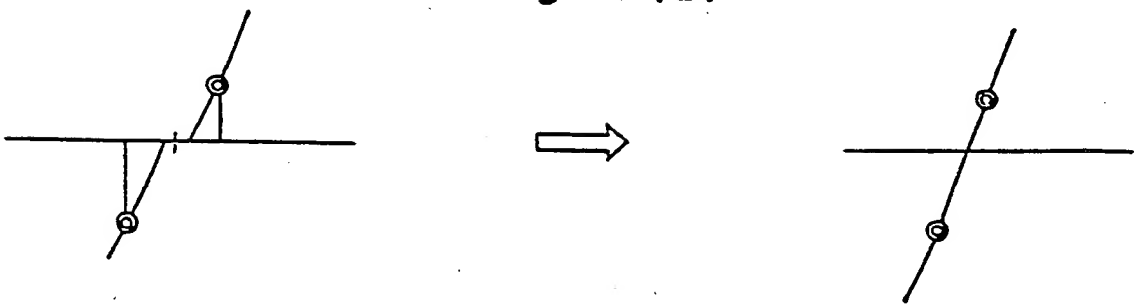


Fig. 7(a)



Verbindungstyp 0b

Fig. 7(b)



Verbindungstyp 1b